Alytes, 1984, 3 (1): 20-24.

SAMPLE-SIZE CONSTRAINTS IN THE USE OF THE NONPARAMETRIC MANN-WHITNEY *U* TEST FOR THE COMPARISON OF TWO INDEPENDENT SAMPLES: CONSEQUENCES IN ANURAN AMPHIBIANS SYSTEMATICS

Alain DUBOIS

Laboratoire des Reptiles et Amphibiens,
Muséum national d'Histoire naturelle,
25 rue Cuvier, 75005 Paris, France

ABSTRACT. - *A table is presented which gives the critical (minimum) value of the size of the larger sample according to that of the smaller sample allowing the use of the Mann-Whitney U test under given conditions (given levels of significance, one-tailed or two-tailed tests). The consequences of these data are outlined, especially as concerns the sample-sizes desirable in fields where the nonparametric U test may be a very useful tool, such as the systematics of the Anuran Amphibians.*

In many biological works, the data from two independent samples are compared statistically. The usual technique for such a work is to apply a Student *t* test to the means of the two groups. However, the correct use of this parametric test requires that certain conditions be satisfied, and among them that the observations be drawn from normally distributed populations having equal variances, and be measured on at least an interval scale (SIEGEL, 1956). Although these conditions are often, at least implicitly (since they are seldom even discussed), assumed to be met with, such assumptions may be unrealistic for some biological data. Even when they are likely to be satisfied, the biologist may prefer to avoid making the assumptions and thus give his conclusions greater generality. In such cases, it is possible to use a nonparametric statistical test. One of the most useful nonparametric tests for the comparison of two independent samples is the Mann-Whitney *U* test, which is almost as powerful as the *t* test and which does not

have the restrictive assumptions and requirements associated with the *t* test, the only requirement of the *U* test being that measurement be achieved at least in an ordinal or ranking scale (SIEGEL, 1956).

In systematic studies of Anuran Amphibians, the Mann-Whitney *U* test has been used for comparisons of morphometrical characteristics of two series of specimens, and especially of series of ratios of two measurements taken from the specimens (INGER, 1966; DUBOIS, 1976, 1983; DUBOIS & KHAN, 1980). The use of a nonparametric test seems here particularly suitable both because of the peculiar properties of ratios (SIMPSON, ROE & LEWONTIN, 1960) and of the special difficulties associated with body measurements of Anurans (DUBOIS, 1977), which are such as to raise doubts as to whether the level of measurement attained is indeed higher than the ordinal scale. In studies of this kind, just as in most other biological works, the levels of significance used are .05, .01 and .001. Finally, in most of the cases no *a priori* hypothesis is tested when comparing two samples, and two-tailed tests must be used (SIMPSON, ROE & LEWONTIN, 1960), but in some cases one-tailed tests may also be called upon: for example, since, in many species of Anurans, females are known to be of a significantly greater size than males (CRUMP, 1974), the validity of this "law" in other species could be tested by means of a one-tailed test.

In the usual conditions specified above (level of significance chosen, two-tailed test), the comparison of two samples by means of the *U* test will not be possible if the size of one of them, or of both, is too small, for in this case the probability associated with the smallest possible value of *U* would be higher than the level of significance retained. To give one example, we may consider two samples of respectively 2 and 5 specimens. In the most extreme case, i.e. when the ranges of the measurements of the two samples do not overlap at all, *U* will take the lowest value possible ($U = 0$). If a double-tailed test is used, the probability associated with $U = 0$ for $n_1 = 2$ and $n_2 = 5$ will be $P = .094$, according to the tables given by SIEGEL (1956): such a value is greater than even the first of our levels of significance (.05), and would be so even if the two samples were drawn from two populations having strongly different characteristics; it is therefore clear that the *U* test cannot be used for comparisons by a double-tailed test of samples having these sizes. The situation would be different, however, if a one-tailed test was involved: the probability associated with $U = 0$ for $n_1 = 2$ and $n_2 = 5$ would then be $P = .047$, i.e. a result indicating

a significant difference at the .05 level. It thus appears that certain combinations of sample-sizes which allow comparisons by one-tailed U tests do not allow them by two-tailed U tests. For a given kind of test, similar differences also exist according to the level of significance chosen.

In systematic studies, especially those dealing with Museums collections, one is often unable to obtain large series of specimens of the same sex and coming from a single population: in many cases the available series consist of less than ten specimens. If an analysis by the U test of the data recorded on this material is intended, it is necessary first to establish whether the sample-sizes available allow the use of this test. As has been discussed above, this will depend not only on the samples-sizes, but also on the level of significance chosen and on the type of test used. In any given case, it is possible to establish, by use of the tables given by SIEGEL (1956), whether a given couple of samples-sizes allows or not the use of the U test, but this may be a rather long procedure. It has therefore appeared useful to prepare a Table giving the critical values of n_1 and n_2 allowing the use of the U test for both two-tailed and one-tailed tests and for the three usual levels of significance (Table I).

In a few cases, for the smallest values of n_1, no comparison is ever possible: even for most unrealistic values of n_2 as high as 10^7, the probability associated with U = 0 is higher than the level of significance. For higher values of n_1, there exists a critical (minimum) value of n_2 allowing comparisons to be made. When n_1 increases, the critical value of n_2 decreases, until both values are equal, a situation which is reached more or less rapidly according to the type of test and the level of significance. While the Table I may be used for ascertaining whether samples of given sizes (from Museums collections for example) may be used for certain types of U tests, it also has the interest of suggesting some rules for the collecting of fresh samples, when the use of the U test may be foreseen for their study. For example, since this test seems to be of a valuable help for morphometrical works dealing with Anurans, it would appear justified to suggest to the workers collecting frogs for such studies to secure at least 8 specimens of each sex from each locality, so as to allow all kinds of uses of the U test in the future on the measurements taken from these specimens. On the other hand, the Mann-Whitney U test being very powerful (SIEGEL, 1956), it does not seem necessary to increase the sample-size very much above this minimum. Since for all values of n_1 and n_2 lower than, or equal to, 20, the

Table I. - Critical (minimum) size of the larger sample (n_2), according to the size of the smaller sample (n_1), allowing a comparison of both samples by the Mann-Whitney U test, in the case of two-tailed and of one-tailed tests and for the levels of significance α of .05, .01 and .001. The symbol NC (no comparison) indicates that, whatever the size of n_2, both samples may not be compared by the U test.

	Two-tailed tests			One-tailed tests		
n_1	α = .05	α = .01	α = .001	α = .05	α = .01	α = .001
1	NC	NC	NC	19	NC	NC
2	8	NC	NC	5	13	NC
3	5	12	NC	3	7	17
4	4	6	49	4	5	10
5	5	5	16	5	5	7
6	6	6	11	6	6	6
7	7	7	9	7	7	7
8	8	8	8	8	8	8

probabilities associated with the values of U have been tabulated (SIEGEL, 1956) and can be found at once without having to calculate Z, it appears appropriate to choose an upper limit of 20 for the size of both samples to be compared. Therefore, for studies using such techniques, a sample-size of 8-20 specimens per sex per locality seems a good standard. In many Anuran species and populations, such quantities of animals may be collected without reducing significantly the population size, and thus without any danger for the population. In the case of small or endangered populations, the collection of specimens should be stopped when the lowest useful values (8 specimens of both sexes) are reached, or even before: in some cases no collection at all should be made.

RESUME

Un Tableau est présenté qui donne la valeur critique (minimale) de

la taille du plus grand échantillon en fonction de celle du plus petit échantillon permettant de comparer ceux-ci à l'aide du test non-paramétrique U de Mann-Whitney dans des conditions données (coefficient de risque choisi: 5 %, 1 % ou 0,1 %; test bilatéral ou unilatéral). Les conséquences de ces données sont soulignées, notamment en ce qui concerne les tailles d'échantillons désirables dans des domaines où l'emploi du test U peut être très utile, comme dans les travaux portant sur la morphométrie et la systématique des Amphibiens Anoures: il est ainsi suggéré, lorsqu'une telle étude est envisagée pour l'avenir, et que cela est possible sans mettre en danger les populations concernées, de récolter au minimum 8 et au maximum 20 animaux de chaque sexe par population étudiée.

REFERENCES BIBLIOGRAPHIQUES

CRUMP, M. L., 1974. - Reproductive strategies in a tropical Anuran community. *Univ. Kansas Mus. nat. Hist. misc. Publ.*, *61*: 1-68.

DUBOIS, A., 1976. - Les Grenouilles du sous-genre *Paa* du Népal (famille Ranidae, genre *Rana*. *Cahiers népalais - Documents*, Paris, C.N.R.S., *6*: i-vi + 1-275.

----- 1977. - Les problèmes de l'espèce chez les Amphibiens Anoures. *Mém. Soc. zool. Fr.*, *39*: 161-284.

----- 1983. - Note préliminaire sur le groupe de *Rana* (*Tomopterna*) *breviceps* Schneider, 1799 (Amphibiens, Anoures), avec diagnose d'une sous-espèce nouvelle de Ceylan. *Alytes*, *2*: 163-170.

DUBOIS, A. & KHAN, M. S., 1980. - A new species of frog (genus *Rana*, subgenus *Paa*) from Northern Pakistan (Amphibia, Anura). *J. Herpet.*, *13*: 403-410.

INGER, R. F., 1966. - The systematics and zoogeography of the Amphibia of Borneo. *Fieldiana: Zool.*, *52*: 1-402.

SIEGEL, S., 1956. - *Nonparametric statistics for the behavioral sciences*. New York, Toronto, London & Tokyo, McGraw-Hill & Kogakusha: i-xvii + 1-312.

SIMPSON, G. G., ROE, A. & LEWONTIN, R. C., 1960. - *Quantitative zoology*. Revised edition. New York & Burlingame, Harcourt, Brace & Co.: i-vii + 1-425.